

# The Heuristic Conception of Inference to the Best Explanation

Finnur Dellsén

Penultimate draft, accepted for publication in *Philosophical Studies*

## Abstract

An influential suggestion about the relationship between Bayesianism and Inference to the Best Explanation (IBE) holds that IBE functions as a heuristic to approximate Bayesian reasoning. While this view promises to unify Bayesianism and IBE in a very attractive manner, important elements of the view have not yet been spelled out in detail. I present and argue for a heuristic conception of IBE on which IBE serves primarily to locate the most probable available explanatory hypothesis to serve as a working hypothesis in an agent's further investigations. Along the way, I criticize what I consider to be an overly ambitious conception of the heuristic role of IBE, according to which IBE serves as a guide to absolute probability values. My own conception, by contrast, requires only that IBE can function as a guide to the comparative probability values of available hypotheses. This is shown to be a much more realistic role for IBE given the nature and limitations of the explanatory considerations with which IBE operates.

## 1 INTRODUCTION

Much of the recent literature on Inference to the Best Explanation (IBE) concerns how, if at all, IBE fits into a Bayesian approach to non-deductive reasoning. Van Fraassen (1989) famously argued that IBE is incompatible with Bayesianism since IBE would, according to van Fraassen, require agents to assign higher probabilities to more explanatory hypotheses than the Bayesian

rule of conditionalization dictates. However, few philosophers have been convinced that van Fraassen's suggestion is the most plausible way of finding a place for IBE in the Bayesian framework (Kvanvig, 1994; Harman, 1997; Douven, 1999). In addition, van Fraassen's conclusion leaves Bayesians in the arguably awkward position of having to deny that explanatory considerations play any substantive role in non-deductive reasoning. While van Fraassen seems to endorse this conclusion, most other philosophers will find it unpalatable given the frequent appeal to explanatory considerations in both scientific and everyday reasoning.

Those who do think, contra van Fraassen, that IBE is compatible with Bayesianism have advocated at least three different approaches to fitting IBE into the Bayesian framework. Some have argued that IBE and Bayesianism are different ways of describing what is essentially the same form of reasoning (Niiniluoto, 1999; Henderson, 2014). Others have argued that explanatory considerations help to determine the objectively correct prior probabilities from which Bayesian agents ought to reason (Weisberg, 2009; Huemer, 2009). Finally, the third and arguably most influential approach has focused on the idea that IBE might serve as a heuristic to approximate correct Bayesian reasoning. On this approach, IBE complements Bayesianism by providing a rule of inference that is appropriate for non-ideal agents and yet enables these agents to approximate the probabilities that Bayesian reasoning would have them assign to hypotheses (Okasha, 2000; McGrew, 2003; Lipton, 2004).

The aim of this paper is to explore the limits of this heuristic approach to IBE and argue for a particular heuristic conception of IBE that respects these limits. More precisely, I present and argue for a heuristic conception of IBE on which IBE serves primarily to locate the most probable available explanatory hypothesis to serve as a working hypothesis in an agent's further investigations. Along the way, I criticize what I consider to be an overly ambitious conception of the heuristic role of IBE, according to which IBE serves as a guide to absolute probability values. My own conception, by contrast, requires only that IBE can function as a guide to the comparative probability values of available hypotheses. This is shown to be a much more realistic role for IBE given the nature and limitations of the explanatory considerations with which IBE operates. While this by no means provides a full account of the relationship

between IBE and Bayesianism, I hope to show that it points in the direction of a more promising heuristic conception of IBE.

I will proceed as follows. In section 2, I start by clarifying what exactly a heuristic conception of IBE would be, and what kind of Bayesian probabilities it is plausibly seen as providing a heuristic for. In section 3, I go on to argue that since IBE involves comparisons between available explanatory hypotheses, explanatory considerations cannot generally function as a reliable heuristic for approximating absolute Bayesian probabilities. In light of this difficulty, I suggest in section 4 that IBE is plausibly construed as a heuristic primarily for comparative probabilities, and argue that this still leaves IBE with an important role to play in a broadly-speaking Bayesian framework, viz. as the basis for choices about which hypotheses to adopt as working hypotheses. Section 5 is the conclusion.

## 2 IBE AS A BAYESIAN HEURISTIC

As it is standardly conceived, Inference to the Best Explanation (IBE) is a rule of inference in which one infers a hypothesis on the grounds that it would, if true, provide a better explanation of one's evidence than any other available alternative hypothesis (Harman, 1965; Thagard, 1978; Lycan, 1988; Lipton, 2004). What makes one explanation better than another is seldom spelled out in detail, but is generally taken to depend on various factors known as *explanatory considerations*. These include:

*Explanatory power*: Other things being equal,  $H_1$  should be preferred to  $H_2$  if  $H_1$  explains more facts (or more kind of facts) than  $H_2$ .

*Antecedent plausibility*: Other things being equal,  $H_1$  should be preferred to  $H_2$  if  $H_1$  fits better than  $H_2$  with what one already has reason to believe.

Numerous other putative explanatory considerations have been proposed – e.g. *simplicity*, *fecundity*, *testability*, *avoidance of ad hoc elements*, and *explanatory depth* – but it's a highly contested issue which (if any) of these other factors IBE should operate with. For example, while *simplicity* is often referred to as

an explanatory virtue, many philosophers of science are quite skeptical of its epistemic force (e.g. Sober, 1990; Dupre, 2002). Since this issue is orthogonal to my concern in this paper, I will seek to avoid any such contested putative explanatory considerations by operating only with the two considerations listed above.

The most detailed and widely discussed account of IBE is Peter Lipton's (2004). On Lipton's view, explanatory considerations are both used to choose between available hypotheses, and to generate the hypotheses that one chooses between in an IBE. So, on Lipton's view, IBE is really a two-step process, where one first generates a limited set of competing explanatory hypotheses and then infers the best of those that have been generated in this way. Lipton qualifies this by saying that in IBE the inferred hypothesis must not only provide the best explanation; this explanation must also be "good enough". Finally, Lipton distinguishes between *likely* and *lovely* explanations, arguing that IBE should be construed as inferring the loveliest explanation, where loveliness is determined by explanatory considerations such as those mentioned above. However, Lipton also claims that loveliness tracks "likeliness", i.e. that lovelier explanations are generally more likely to be true. Since Bayesian reasoning concerns the probability of hypotheses, Lipton's contention that loveliness tracks likeliness underwrites the idea that IBE may provide a heuristic for Bayesian reasoning.

As Lipton (2004, 107) himself recognizes, the philosophical project of articulating how IBE would provide such a heuristic for Bayesian reasoning is still in its infancy. One question – examined below – is what kind of Bayesian reasoning IBE should be seen as providing a heuristic for. Before we delve into that issue, however, we must say something about what it would be for IBE to be a *heuristic* to Bayesian reasoning instead of the fundamental and free-standing inference rule that it is perceived to be by many other authors (see, e.g., Harman, 1965; Foster, 1982; Lycan, 1988, 2012; Psillos, 2007; Weintraub, 2013). The idea, in short, is that IBE provides a procedure that is both usable by ordinary epistemic agents and yet approximates Bayesian reasoning. IBE would thus provide a kind of guideline or recipe that the agent may follow in her doxastic deliberation – whereas Bayesianism provides the ultimate standard of evaluation of such doxastic attitudes. Of course, IBE can only serve as

a heuristic in this way if reasoning in accordance with IBE is something that does in fact approximate Bayesian reasoning (an issue to which we will return below). That said, IBE need not accord with Bayesian reasoning exactly and in every case in order to serve as a heuristic in this sense, since – just as with any other heuristic – it may fail in extraordinary cases and provide only a rough guide in the cases in which it does not fail.<sup>1</sup>

This heuristic conception of IBE assumes that the kind of explanatory reasoning involved in IBE is more accessible to ordinary human agents than the kind of probabilistic reasoning required by Bayesianism. There is considerable empirical evidence for this assumption. Well-known results from Tversky and Kahneman (1984) have inspired extensive research into the various ways in which ordinary people are quite poor at probabilistic reasoning. From a Bayesian point of view, there is thus a clear advantage to having accessible heuristics to approximate Bayesian reasoning, which is what IBE promises to deliver on the heuristic conception. Of course, if IBE is to play this role, ordinary agents must at least be able to, and preferably be disposed to, actually reason in accordance with the recommendations of IBE. Happily, recent empirical research into the role of explanation in reasoning suggests that agents do indeed prefer theories that are more explanatorily powerful (Read and Marcus-Newhall, 1993; Preston and Epley, 2005), fit better with what one already believes (Pennington and Hastie, 1992; Sloman, 1994), and that exhibit a kind of simplicity (Read and Marcus-Newhall, 1993; Lombrozo, 2007). Summarizing this body of work, Lombrozo (2016b, 749) says that “when children and adults generate and evaluate explanations, they recruit explanatory virtues, such as simplicity and breadth, as evaluative constraints on reasoning.”<sup>2</sup>

The other main issue for a heuristic conception of IBE concerns what sort of Bayesian reasoning IBE is meant to provide a heuristic for. In general,

---

<sup>1</sup>I take this characterization of what it would be for IBE to serve as a heuristic for Bayesian reasoning to be in harmony with Chow’s (2015) recent in-depth study of the concept of a “heuristic” as used in cognitive science. Chow characterizes inferential cognitive heuristics as “satisficing cognitive procedures that can be expressed as rules one reasons in accordance with; they require little cognitive resources for their recruitment and execution; they operate by exploiting concepts” (Chow, 2015, 45).

<sup>2</sup>For other summaries and interpretations of this research, see Keil (2006) and Lombrozo (2006, 2010, 2011, 2016a). Much of this empirical work is inspired by Thagard (1989).

Bayesians hold that the credences of rational agents should satisfy the Kolmogorov axioms of probability, and can therefore be represented as probabilities. However, there are many distinct ways to be a Bayesian in this sense, and some Bayesian views may be more plausibly coupled with IBE than others.<sup>3</sup> For our purposes, *Subjective Bayesianism* will be taken to be the view that the only other constraint on the credences of rational agents is the diachronic requirement of *Bayesian Conditionalization*. Of course, Subjective Bayesianism (thus construed) is an extremely permissive view of epistemic rationality, with very minimal constraints requiring only a kind of formal synchronic and diachronic coherence in a rational agent's credences. Accordingly, most epistemologists (including many Bayesians) hold that there are some additional constraints on rational credences, although there is disagreement on what they are and how they should be characterized. These additional constraints might be principles connecting credences to physical chances such as the Principal Principle (Lewis, 1980), principles connecting credences to symmetry considerations such as the Principle of Indifference (Laplace, 1951; Keynes, 1921) and the Maximum Entropy Principle (Jaynes, 2003; Rosenkrantz, 1977), or even principles that constrain rational credences based on explanatory considerations (Huemer, 2009; Weisberg, 2009). Here I will refer to views that place some additional constraints of this sort on rational credences as *objective Bayesian views*, keeping in mind that this category includes a range of views that accept different constraints on rational credences.

So which kind of Bayesianism is IBE most plausibly seen as providing a heuristic for? Jonathan Weisberg (2009) argues that coupling IBE with Subjective Bayesianism would “rob IBE of some of its most interesting applications [and] much of its intuitive appeal” (Weisberg, 2009, 135-136). On Weisberg's view, the basic problem with construing IBE as a heuristic for Subjective Bayesianism is that the two modes of inference track fundamentally different things: IBE tracks explanatory loveliness, while Subjective Bayesianism tracks diachronic consistency with an agent's prior credences. Since an agent's prior credences need not exhibit a preference for more explanatory hypotheses according to Subjective Bayesianism, IBE and Subjective Bayesianism deliver

---

<sup>3</sup>See e.g. Good's (1971) well-known estimation that there are at least 46656 distinct Bayesian views.

conflicting verdicts in many cases. In my view, Weisberg’s criticism is compelling, and so I think he is correct to dismiss a conception of IBE on which IBE serves as a heuristic for Subjective Bayesianism. However, it’s important to note that it does not follow from Weisberg’s criticism that IBE could not function as a heuristic for *any* Bayesian reasoning. Specifically, Weisberg’s criticism leaves open that IBE may serve as a heuristic to objective Bayesian views, which posit constraints upon rational credences that go beyond those of synchronic and diachronic probabilistic consistency. Indeed, Weisberg himself suggests that IBE should be seen as providing objectivist constraints on rational probability assignments. A similar suggestion is made by Leah Henderson (2014), who argues that the preference for explanatory hypotheses follows from standard objectivist constraints such as the Maximum Entropy Principle. Either way, the preference for explanatory theories recommended by IBE would also appear in the probability distributions licensed by these objectivist constraints.

There are certainly important remaining questions about how the preference for explanatory theories would either grow out of, or be posited into, objective Bayesian views in this way.<sup>4</sup> I will not be focusing on these issues here, since my concern is with the extent to which IBE could serve as a *heuristic* for objective Bayesian reasoning, i.e. with whether the former may be a doxastic decision procedure that is appropriate for realizing the latter in ordinary agents. Weisberg and Henderson do not address this issue since they are not concerned with arguing that IBE may work as a heuristic in this way. In order for IBE to serve a heuristic role of this kind, it is not enough that IBE and objective Bayesian constraints substantially agree on how an agent ought to go about her non-deductive reasoning; IBE must also provide a relatively accessible way for ordinary agents to make doxastic decisions in light of the available evidence. On the face of it, IBE does seem well-suited for playing this role, since the explanatory considerations to which it appeals certainly appear more accessible than the often elusive probability constraints and calculations

---

<sup>4</sup>I am myself most sympathetic to a Henderson-style view on which preferences for more explanatory theories grow out of independently-motivated objectivist constraints on probabilities. Henderson refers to this view of the relationship between IBE and Bayesianism as *emergent compatibilism*, since it means that “IBE would ‘emerge’ without the Bayesian doing anything differently from what she would anyway” (Henderson, 2014, 699).

that are the basis of objective Bayesian reasoning.<sup>5</sup> However, as we shall now see, this idea is not as straightforward as one might have thought.

### 3 THE LIMITS OF EXPLANATORY HEURISTICS

So far I have suggested that the most plausible version of the heuristic conception of IBE is one on which IBE functions as a procedure for approximating objective Bayesian reasoning. In this section, I consider a problem for this heuristic conception of IBE, which I will argue should lead us to settle for a more moderate heuristic conception of IBE. The problem, in short, is that IBE involves an essentially *comparative* evaluation of *available* hypotheses, while Bayesian reasoning involves an non-comparative or absolute evaluation of such hypotheses. To draw out the problem, I will first consider what I take to be a relatively unproblematic way in which IBE could serve its heuristic role, viz. as a guide to comparative probabilities, and then illustrate the difficulties involved in having IBE play the more ambitious role of providing a guide to absolute probabilities.

#### 3.1 IBE AND COMPARATIVE PROBABILISTIC EVALUATIONS

Suppose we have some evidence  $E$ , background knowledge  $B$ , and a set  $\mathbf{H}_A = \{H_1, \dots, H_n\}$  of available hypotheses that provide competing potential explanations of  $E$  (perhaps in conjunction with auxiliary assumptions from  $B$ ). On an objectivist heuristic conception, IBE provides a heuristic guide to the objectively correct probabilities that should be assigned to these hypotheses in light of  $E$  and  $B$ . The objectivist Bayesian accepts a version of Bayesian Conditionalization according to which the credence you should assign to a hypothesis  $H_i \in \mathbf{H}_A$ , having obtained evidence  $E$ , should equal the credence that you should have assigned, before obtaining  $E$ , to  $H_i$  conditional on  $E$ . So, according to objectivist Bayesian views, one's credence in  $H_i$  given  $E$  and background knowledge  $B$  should be  $P(H_i|E \wedge B)$ , where  $P(\cdot)$  is a probability function sanctioned by the objectivist principles or constraints in question.

---

<sup>5</sup>From now on, I will use “Bayesian” as short for “objective Bayesian” unless otherwise indicated.



On the heuristic conception we are currently considering, IBE functions as a guide to assigning probabilities of this sort.

Note that by Bayes's Theorem, the *comparative* posterior probabilities of two hypotheses are determined by their comparative priors and likelihoods. That is, an inequality such as

$$P(H_i|E \wedge B) > P(H_j|E \wedge B) \quad (1)$$

is, by Bayes's Theorem, equivalent to:

$$P(H_i|B) P(E|H_i \wedge B) > P(H_j|B) P(E|H_j \wedge B) \quad (2)$$

Now, several authors have argued that explanatory considerations track priors and likelihoods. In particular, it has been argued that, all other things being equal,  $P(E|H_i \wedge B) > P(E|H_j \wedge B)$  whenever  $H_i$  has greater explanatory power than  $H_j$  vis-à-vis  $E$  (McGrew, 2003; Henderson, 2014). Similarly, it seems plausible that, all other things being equal, a hypothesis that fits better with what one already has reasons to believe is more probable irrespective of the new evidence  $E$ , and thus that  $P(H_i|B) > P(H_j|B)$  whenever  $H_i$  is antecedently more plausible than  $H_j$  in the relevant sense (Okasha, 2000; Lipton, 2004). As noted above, I won't be assessing these claims here, since my concern is not with the claim that explanatory considerations track objective Bayesian probabilities *per se*, but with the idea that the explanatory considerations may function as a heuristic for objective Bayesian reasoning.

Assuming that there is, at least *ceteris paribus*, such a robust connection between the relevant set of explanatory considerations on the one hand and priors and likelihoods on the other hand, it does appear quite plausible that IBE may serve its heuristic function when it comes to *comparing* the probabilities of two or more competing explanatory hypotheses. This is best illustrated by an example: Suppose a detective arrives at a crime scene, where a woman lies dead from a shot wound. Her brother called in to report the crime, claiming that he had found his sister in this state when he arrived at her home. However, searching the apartment, no suicide note was found. Let these pieces of information be the relevant evidence  $E$  in this case. Having considered this

evidence for just a moment, our detective forms three hypotheses concerning the victim's cause of death:

$H_b$ : The victim's brother shot her.

$H_s$ : The victim shot herself.

$H_a$ : The victim's brother hired an assassin to shoot her.

For the detective, these are the available competing explanatory hypotheses vis-à-vis the detective's evidence  $E$ . Note that this set of available hypotheses is not exhaustive: the actual cause of death might be something else entirely.<sup>6</sup>

Now let's suppose that the detective reasons as follows. Comparing  $H_b$  and  $H_s$ , the detective notes that  $H_b$ , in contrast to  $H_s$ , provides an explanation for the fact that there was no suicide note; and since  $H_b$  and  $H_s$  are at least roughly equal with regard to other explanatory considerations, she concludes that explanatory considerations favor  $H_b$  over  $H_s$ . This reasoning tracks what may plausibly be taken as the comparative objective probabilities of  $H_b$  and  $H_s$ , since the fact that  $H_b$  explains more of the evidence  $E$  than  $H_s$  may be taken to indicate that  $P(E|H_b \wedge B) \gg P(E|H_s \wedge B)$ , while the fact that  $H_s$  and  $H_b$  are roughly equal in other explanatory considerations indicates that  $P(H_b|B) \approx P(H_s|B)$ . It then follows, from the equivalence of (1) and (2), that  $P(H_b|E \wedge B) > P(H_s|E \wedge B)$ .

A similar point applies to a comparison of  $H_b$  and  $H_a$ . Any detective will know beforehand (i.e. irrespective of  $E$ ) that hired assassinations are extremely rare; indeed, they will know that murders are almost always committed by the victims' family members or acquaintances. The detective would thus conclude that  $H_b$  is antecedently more plausible than  $H_a$ , and since  $H_b$  and  $H_a$  are at least roughly equal with regard to other explanatory considerations (such as explanatory power), she concludes that explanatory considerations favor  $H_b$  over  $H_a$ . Again, this reasoning plausibly tracks comparative objective probabilities, since the fact that  $H_b$  is antecedently more plausible than  $H_a$  indicates that  $P(H_b|B) \gg P(H_a|B)$ , and the fact that  $H_b$  and  $H_a$  are roughly equal with regards to other explanatory considera-

---

<sup>6</sup>Indeed, available hypotheses almost never are exhaustive. This is an important point to which will return below.

tions indicates that  $P(E|H_b \wedge B) \approx P(E|H_a \wedge B)$ , from which it follows that  $P(H_b|E \wedge B) > P(H_a|E \wedge B)$  by the same token as before.

The important point here is that the detective's reasoning in terms of explanatory considerations plausibly delivers the same preference for  $H_b$  over  $H_s$  and  $H_a$  as does a comparison of objective Bayesian probabilities, and yet reasoning in terms of explanatory considerations is considerably more tractable for ordinary agents such as the detective. Plausibly, the detective can tell almost immediately that  $H_b$  is preferable to  $H_s$  in light of its greater explanatory power, while she may not be able to reason probabilistically to the same conclusion nearly as quickly – or even at all. Similarly, the detective can tell immediately that  $H_b$  is preferable to  $H_s$  in light of its being antecedently more plausible, while she cannot immediately determine the relative probabilities as quickly – or even at all – by calculating the priors and likelihoods of these hypotheses. In sum, then, IBE does appear to be able to serve a heuristic role for objective Bayesian reasoning in that it provides an accessible way of estimating comparative probabilities of explanatory hypotheses relative to a given set of evidence.

So far I have focused on cases in which some explanatory considerations favor one hypothesis over another while other explanatory considerations remain roughly neutral on the two hypotheses. Of course, there will also be cases in which explanatory considerations pull in opposite directions. As a case in point, consider a comparison between  $H_s$  and  $H_a$ .  $H_s$  is antecedently more plausible than  $H_a$  relative to  $B$ , but  $H_a$  explains more of  $E$  since it explains the fact that the victim did not leave a suicide note. In this particular comparison, we may well want to say that the comparison in explanatory power is so decisive as to clearly outweigh the fact that antecedent plausibility favors  $H_s$  over  $H_a$ . On the probabilistic side, this would mean that  $P(E|H_s \wedge B) \ll P(E|H_a \wedge B)$  while  $P(H_s|B) \gg P(H_a|B)$ , so that  $P(H_s|E \wedge B) < P(H_a|E \wedge B)$ . However, in other cases, the comparisons might be closer, with IBE offering no clear verdict on which hypothesis should be preferred. In such cases the probabilities of the two hypotheses are simply too close for IBE to make a decisive call. This is what we should expect from a heuristic form of reasoning, since heuristics are designed only to give approximate judgments and may be of limited use when greater precision is called for.

At any rate, it's worth highlighting that if and in so far as IBE provides a heuristic to compare the probabilities of two hypotheses in this way, it also provides a heuristic for ranking a set of three or more hypotheses from most to least probable, allowing for ties and comparisons that are too close to call. Notice that this ranking is obtained by comparative probabilities only – no absolute probabilities need to be assigned in order to obtain the ranking. For example, in order to rank  $H_b$  as most probable among  $H_b$ ,  $H_s$  and  $H_a$ , we did not need to estimate the absolute (i.e. non-comparative) values for their priors  $P(H_b|B)$ ,  $P(H_s|B)$  and  $P(H_a|B)$ . Thus, in so far as IBE can provide a reliable heuristic guide to comparative probabilities, it can also provide such a guide for probability rankings.<sup>7</sup> In particular, then, IBE could serve as a heuristic for identifying the hypotheses with the highest probabilities in a given set of available explanatory hypotheses. (This point will be important below.)

### 3.2 IBE AND ABSOLUTE PROBABILISTIC EVALUATIONS

So far we have examined how IBE may provide a heuristic for making comparative judgments about the objective Bayesian probabilities of two or more explanatory hypotheses. However, any standard version of Bayesianism operates with more than comparisons between probabilities of two or more hypotheses. Bayesians are interested not only in whether  $H_i$  is more probable than  $H_j$ , but also how probable each of  $H_i$  and  $H_j$  are in light of the evidence  $E$  at a given time. In principle, these absolute probabilities can of course be calculated in the usual way by appealing to a standard version of Bayes's Theorem:

$$P(H_i|E \wedge B) = \frac{P(H_i|B) P(E|H_i \wedge B)}{P(E|B)} \quad (3)$$

The question, however, is whether IBE helps ordinary agents do this with any reliability, such that we can say that IBE provides an accessible heuristic

---

<sup>7</sup>If this seems puzzling, consider an analogy: Suppose you needed to find the tallest person in a room without measuring anyone's height. One way to do this is to ask everyone to line up according to the following rule: Make sure the person on your left is at least as short as you are and that the person on your right is at least as tall as you are. Assuming everyone follows the rule, the people in the room will now be ordered from smallest to tallest without determining anyone's absolute height.

for objective Bayesian reasoning with absolute probability values. I will now argue that IBE does not provide us with a heuristic for this purpose in typical cases, since the explanatory considerations appealed to in IBE are generally not suitable for indicating absolute probability values.

The problem that I will be focusing on can be brought out by noticing a crucial difference between absolute and comparative probabilistic evaluations. By (3), the value of  $P(H_i|E \wedge B)$  depends crucially on the *marginal likelihood*  $P(E|B)$  – a term that drops out when the probability of two hypotheses are being compared (as in (1) and (2)). Unlike the prior  $P(H_i|B)$  and likelihood  $P(E|H_i \wedge B)$ , the marginal likelihood  $P(E|B)$  does not map neatly onto any explanatory considerations possessed by  $H_i$  vis-à-vis  $E$ , since these are all either features of the  $H_i$ ’s relationship with the background knowledge  $B$  (e.g. antecedent plausibility) or  $H_i$ ’s relationship with the evidence  $E$  (e.g. explanatory power). Indeed, since  $H_i$  does not even occur in  $P(E|B)$ , it seems clear that the value of this probability is not tracked by any explanatory consideration of  $H_i$  in any straightforward way (whether this probability is tracked by explanatory considerations in some indirect way will be discussed below). Perhaps for this reason, those who argue that IBE can be located in the Bayesian framework tend to focus on priors and likelihoods, and leave out discussions of how explanatory considerations would track the marginal likelihood  $P(E|B)$  (see, e.g., Okasha, 2000; Weisberg, 2009; Henderson, 2014).

To my knowledge, Lipton (2004, 115-116) is the only proponent of the heuristic conception of IBE who makes any suggestion for how IBE might track the marginal likelihood. Echoing a common Bayesian sentiment, Lipton refers to  $P(E|B)$  as “tantamount to how surprising it would be to observe  $E$ ” (Lipton, 2004, 116).<sup>8</sup> If  $P(E|B)$  is to be constrained in line with objective Bayesian views (see §2.2), this must be taken to refer to how surprising it *should* be to observe  $E$ , as opposed to how surprised an agent would in fact be if she were to observe  $E$ . Indeed, Lipton immediately goes on to say that IBE helps us track  $E$ ’s surprisingness because it “will be determined in part

---

<sup>8</sup>This is slightly misleading since it may seem to suggest that surprising evidence has a *high* marginal likelihood. Presumably, the quantity Lipton meant to be describing is not the marginal likelihood  $P(E|B)$  but its reciprocal  $\frac{1}{P(E|B)}$  (which is what gets multiplied with the prior and likelihood in Bayes’s Theorem).

by how good an explanation of  $E$  my current beliefs would supply.” (Lipton, 2004, 116)

Let’s try to unpack Lipton’s idea here. In a Bayesian framework, an agent’s “current beliefs” are her credences prior to conditionalization; and according to objective Bayesian views, such credences should equal the objective Bayesian probabilities conditional on background evidence. Some of these probabilities will be assigned to hypotheses that would, if true, explain the evidence  $E$  – call these hypotheses  $H_1$ - $H_m$ . If we assume that  $H_1$ - $H_m$  are mutually exclusive and jointly exhaustive,<sup>9</sup> we can write  $P(E|B)$  as a function of the priors and likelihoods of  $H_1$ - $H_m$  thus:

$$P(E|B) = \sum_{j=1}^m P(H_j|B) P(E|H_j \wedge B) \quad (4)$$

So if explanatory considerations track the priors and likelihoods of each hypothesis  $H_1$ - $H_m$ ,  $P(E|B)$  would indeed be determined by the explanatory loveliness of  $H_1$ - $H_m$  vis-à-vis the evidence  $E$ , as Lipton’s comment suggests.

Although this idea is quite attractive from a purely formal point of view, the question that is relevant for our purposes is whether estimating explanatory loveliness of each hypothesis  $H_1$ - $H_m$  is an accessible heuristic for tracking  $P(E|B)$  and thus for indicating the absolute probability of  $H_i$  given  $E$ . One obvious problem is that it would require agents using IBE as a heuristic to have some reliable way of simultaneously estimating the explanatory loveliness of each and every one of the hypotheses  $H_1$ - $H_m$  before making an inference by IBE. In addition, such agents would presumably have to have some way of aggregating the explanatory loveliness of each hypothesis into an overall

---

<sup>9</sup>It might seem unreasonable that  $H_1$ - $H_m$  could be jointly exhaustive, since one possibility that cannot be ruled out *a priori* (and thus should not be assigned probability 0) is that there is no explanation for  $E$ . However, there are views on which the claim that there is no explanation for a phenomenon counts as a (degenerate) explanation of that phenomenon. For example, Lewis (1986) and Skow (2014) both hold that the information that an event is uncaused or coincidental counts as an explanation of that event. So, for the sake of Lipton’s argument, I will assume that  $H_1$ - $H_m$  are indeed exhaustive and thus that they include a hypotheses on which there is no explanation for  $E$ . (See also (Dellsén, 2016a) for an argument that competing explanatory hypotheses should be mutually exclusive and jointly exhaustive in a maximally plausible conception of IBE.)

estimation of the explanatory loveliness provided by them all. This would be a daunting task for any ordinary agent, which casts serious doubts on whether IBE could serve as an accessible heuristic if this were required.<sup>10</sup> I want to set this problem aside, however, since there is an even more serious problem with this proposal.

The problem, in short, is that IBE operates on a limited number of available explanatory hypotheses as opposed to a set of jointly exhaustive hypotheses.<sup>11</sup> When ordinary agents infer by IBE from some evidence  $E$ , they have (at least normally) considered only a small fraction of all the hypotheses  $H_1-H_m$  which provide potential explanations of a given  $E$ , so tracking  $P(E|B)$  by means of estimating the explanatory loveliness of all potential explanations of  $E$  would require agents to estimate the explanatory loveliness of hypotheses they have never even considered. This would clearly make IBE useless as a heuristic. In our detective case from before, we would have to estimate the explanatory loveliness not only of  $H_b$ ,  $H_s$  and  $H_a$  but also of all the potential explanations that the detective has not even considered. In that case, the detective could not employ IBE in the first place – she would be forced to wait until she had considered all possible explanations of the evidence. If that were required for IBE to apply, IBE would clearly be useless in this case.

Of course, we *could* conceive of IBE as selecting the best explanation from a set of jointly exhaustive hypotheses – thus requiring of IBE-employing agents that they have already exhausted the logical space of possible explanations of  $E$ . But then IBE would have no hope of providing an accessible heuristic for ordinary agents in normal situations, since such agents have rarely, if indeed ever, considered all explanatory hypotheses in logical space. This problem can be thought of as a dilemma for using explanatory considerations to track  $P(E|B)$ : Either the available explanatory hypotheses with which IBE operates are required to include all possible explanations of the evidence  $E$ , or they are

---

<sup>10</sup>Of course, it is ultimately an empirical issue whether this is something ordinary agents are capable of doing. But it would surely be more than a little surprising if ordinary agents were capable of this extremely sophisticated sort of explanatory reasoning in light of how poor we are at related cognitive tasks.

<sup>11</sup>This is acknowledged by all proponents of IBE of which I am aware who comment on the issue at all, e.g. Lycan (1988, 129) and Lipton (2004, 149).

not. If they are, then IBE would not be usable by ordinary agents in typical cases, since the agents have not considered all possible explanations of  $E$  in such cases. If they are not, then the estimations of explanatory loveliness of available explanatory hypotheses do not (even if correct) track  $P(E|B)$  in typical cases, since they would determine the priors and likelihoods of only a small fraction of the hypotheses occurring on the right-hand side of (4). Either way, IBE would not provide an accessible heuristic that tracks  $P(E|B)$  in typical cases.<sup>12</sup>

Having said this, I wish to acknowledge that there is a class of cases in which this problem does admit of a qualified solution. Suppose one has good reasons to believe that all but one explanatory hypothesis  $H_i$  for some set of facts, including those that one has not considered, will provide so poor explanations for  $E$  that their contribution to the value of  $P(E|B)$  will be negligible. In that case, notice that all the terms in the right-hand side of (4) except for those in which  $H_i$  occurs will be close to 0. Thus we have:

$$P(E|B) \approx P(H_i|B) P(E|H_i \wedge B) \quad (5)$$

By (3), it then follows that  $P(H_i|E \wedge B) \approx 1$ .<sup>13</sup> So, in cases in which we have reason to believe that all other explanations will be so poor as to be negligible, we have reason to believe that the probability of the explanatory hypothesis in question will be very high (i.e. close to unity). In this way, we may well be

---

<sup>12</sup>The problem that I am discussing here will remind some readers of van Fraassen’s *Bad Lot Objection* against IBE. (van Fraassen, 1989, 142-3) This objection holds that IBE cannot be an inductively strong form of inference since the available explanatory hypotheses from which one is choosing in an IBE might very well not include a true explanatory hypothesis; however, if the lot of available explanatory hypothesis does not include a true hypothesis, then IBE will inevitably lead to one inferring a false hypothesis. Although the problem with which I am concerned here is superficially similar to the Bad Lot Objection, the two should not be confused. This is partly because van Fraassen’s Bad Lot Objection is directed at an entirely different conception of IBE – a non-probabilistic and (by van Fraassen’s own lights) unsophisticated version of IBE. More importantly, the problem with which I am concerned is not that the true explanatory hypothesis might not be in the “lot” of available hypotheses we are considering, but that IBE cannot appeal to the explanatory loveliness of the hypotheses that are not included in this “lot”. So the problem I am interested in arises even for “good” lots (i.e. lots that contain a true hypothesis), provided only that these lots do not exhaust the logical space of possible explanations of the evidence.

<sup>13</sup>Of course,  $P(E|B)$  cannot be *higher* than the product of  $P(H_i|B)$  and  $P(E|H_i \wedge B)$ . Nor could  $P(H_i|E \wedge B)$  be higher than 1.



able to use explanatory considerations as a kind of heuristic to estimate that the probability of  $H_i$  will be very high.<sup>14</sup>

While I hesitate to call this an Inference to the *Best* Explanation,<sup>15</sup> this would certainly be a form of reasoning in which explanatory considerations provide a heuristic for estimating absolute probabilities of a certain kind. Of course, it only applies in a circumscribed range of cases, viz. those in which we have good reason to believe that all alternative explanations are so poor as to be negligible. On the other hand, this type of reasoning may well provide an apt description of how scientists arrived at many currently accepted scientific theories. For example, it seems to me that we have good reason to believe that every genuine alternative to the atomic theory of matter will be so poor an explanation of our total set of evidence about the submicroscopic structure of the world that their contribution to the marginal likelihood will be negligible, allowing us to conclude that the probability of the atomic theory is close to unity. This puts me in direct opposition to anti-realists such as Stanford (2006) who are skeptical of inferences of this kind, but that is a place where I am quite happy to be.<sup>16</sup>

So I do not deny that it is possible for explanatory considerations to provide a rough guide to absolute probabilities in the limiting case where we have good reason to believe all other explanations are so poor as to have a negligible effect on the marginal likelihood. However, in typical cases in which IBE is meant to apply – e.g. in our detective case from before – we cannot make this assumption. In those cases, the absolute probability of the hypotheses under consideration will depend crucially on the probability of alternative explanatory hypotheses, most of which we have not even considered in a typical IBE. Since IBE predictably does not provide a heuristic for tracking these probabilities, IBE is unsuitable as a heuristic for estimating the absolute values of objective Bayesian probabilities. Now, as I have indicated, I still think IBE

---

<sup>14</sup>See also (Dellsén, 2017a) for a different way in which we might sometimes be able to use explanatory considerations to estimate that the probability of a hypothesis is high.

<sup>15</sup>“Inference to the Only Explanation” may be a more appropriate term for this sort of inference. For an explanationist account along these lines, see Bird (2005, 2007). See also Earman (1992) for a Bayesian account of eliminative inferences.

<sup>16</sup>And a place where we should all be happy to be, as argued in (Dellsén, 2016b).

has an important heuristic role to play with respect to Bayesian reasoning, even in typical cases – albeit a different and more modest role than one might have otherwise thought. Before I get to that, however, let me consider two possible objections to the argument of this section.

### 3.3 OBJECTIONS AND REPLIES

*Objection 1:* The first objection is based on a logical maneuver discussed in a similar context by Lipton (1993, 94-96).<sup>17</sup> Above I suggested that comparative evaluations between available explanatory hypotheses are relatively unproblematic for IBE. Now, the idea behind the current objection is to reduce any absolute evaluation of a hypothesis  $H_i$  to a comparative evaluation of  $H_i$  and the corresponding “null hypothesis”,  $\neg H_i$ . Notice that  $P(H_i|E \wedge B) > 0.5$  just in case  $P(H_i|E \wedge B) > P(\neg H_i|E \wedge B)$ , which in turn holds if and only if:

$$P(H_i|B) P(E|H_i \wedge B) > P(\neg H_i|B) P(E|\neg H_i \wedge B) \quad (6)$$

If IBE provides a reliable heuristic for comparative evaluations such as these, then it would also provide a reliable heuristic for a kind of absolute evaluation.<sup>18</sup> In this way, comparative evaluations would at least take us a long way towards providing estimations of absolute probabilities as well.

*Reply:* I’m happy to concede that there may be some cases in which the absolute probabilities may be approximated in this way, viz. when  $H_i$  and  $\neg H_i$  both provide potential explanations of the evidence at hand.<sup>19</sup> Generally, however, if  $H_i$  provides a potential explanation of  $E$ , its negation  $\neg H_i$  does not. As a consequence, its explanatory power is trivially nonexistent – or, if you

---

<sup>17</sup>Interestingly, a suggestion along these was already discussed and rejected by van Fraassen (1980, 21-22).

<sup>18</sup>Indeed, by the same token,  $P(H_i|E \wedge B) \approx 1$  just in case  $P(H_i|E \wedge B) \gg P(\neg H_i|E \wedge B)$ , which holds if and only if:

$$P(H_i|B) P(E|H_i \wedge B) \gg P(\neg H_i|B) P(E|\neg H_i \wedge B) \quad (7)$$

So, if all explanatory considerations strongly favor one hypothesis over another, then we could perhaps use IBE to estimate that  $P(H_i|E \wedge B) \approx 1$ .

<sup>19</sup>I have not been able to think up any cases of this sort, but I am also not aware of any reason why such cases would be impossible in principle.

prefer, undefined. To illustrate, consider  $H_b$  from our detective example. Here,  $\neg H_b$  is the hypothesis that the victim's brother did not shoot his sister. While this hypothesis is certainly intelligible, notice that it provides no explanation at all of the evidence at hand, since it doesn't tell you anything about the victim's cause of death apart from the negative claim that  $H_b$  is false. In so far as it makes sense to compare the explanatory power of  $\neg H_b$  to that of  $H_b$  at all, this consideration would thus trivially favor  $H_b$  over  $\neg H_b$  in virtue of  $\neg H_b$ 's inability to provide *any* explanation of the victim's death.

To see why this kind of trivial favoring of a hypothesis over its negation cannot be correct (at least not in an objective Bayesian framework), consider again a set of mutually exclusive and jointly exhaustive explanatory hypotheses  $H_1$ - $H_m$ . If comparisons of explanatory power between each of these hypotheses and their negations are allowed (and supposing that such comparisons make sense at all), they would trivially favor each explanatory hypothesis  $H_j$  over its null hypothesis  $\neg H_j$  (where  $1 \leq j \leq m$ ). Since explanatory power is meant to track likelihoods, we would then get:

$$P(E|H_j \wedge B) > P(E|\neg H_j \wedge B), \text{ for } 1 \leq j \leq m \quad (8)$$

This amounts to saying that the likelihood ratio (relative to  $E$ ) is positive for all these hypotheses  $H_1$ - $H_m$ , which in turn entails that conditionalizing on  $E$  raises the probability of every single one of these hypotheses:

$$P(H_j|E \wedge B) > P(H_j|B), \text{ for } 1 \leq j \leq m \quad (9)$$

However, since  $H_1$ - $H_m$  are mutually exclusive and jointly exhaustive, the probability axioms dictate that both their prior probabilities, and their posterior probabilities, must sum to one:

$$\sum_{j=1}^m P(H_j|B) = 1 \quad (10)$$

$$\sum_{j=1}^m P(H_j|E \wedge B) = 1 \quad (11)$$

Hence, contra (9) (and (8)), it is impossible for  $E$  to raise the probability of

all these hypotheses  $H_1$ - $H_m$  (although it is of course possible for  $E$  to raise the probability of *some* of them provided that the probability of some of the other hypotheses is lowered by  $E$ ). To put the point differently, any agent who trivially favors any explanatory hypothesis over its negation by always assigning higher likelihoods to the former than to the latter will end up being probabilistically incoherent. This clearly would not serve as a reliable guide to objective Bayesian probabilities (or, indeed, to anything else that satisfies the probability axioms).

In sum, then, Lipton’s suggestion that that we evaluate the absolute probability of  $H_i$  by comparing its explanatory loveliness with its corresponding null hypothesis  $\neg H_i$  would lead to a trivial favoring of  $H_i$  over  $\neg H_i$  in terms of their respective explanatory powers. Given that comparisons in explanatory power is meant to correspond to comparisons of likelihoods on the current suggestion, this kind of trivial favoring leads to incoherent probability assignments. To avoid this absurd conclusion, we must reject Lipton’s suggestion that IBE provides any way of evaluating the explanatory loveliness of a null hypothesis  $\neg H_i$  in relation to  $E$ , except possibly in the rare cases in which the negation of an explanatory hypothesis itself provides a potential explanation of  $E$  (assuming such cases exist at all). Since this is not the case in most – and certainly not in all – cases in which IBE is meant to apply, the logical maneuver of obtaining an absolute evaluation of  $H_i$  by means of comparing it with  $\neg H_i$  will not work generally.<sup>20</sup>

*Objection 2:* The second objection is “transcendental”. According to this objection, ordinary agents routinely make estimations of the absolute probabilities of various hypotheses, including hypotheses that are meant to provide explanations. For example, we say that Darwin’s theory of natural selection is very likely to be a true, while Lamarck’s theory of acquired characteristics is likely to be false. These are estimations of absolute probabilities as opposed to merely comparative estimations. If my arguments are correct, however, IBE would not provide a reliable heuristic for this purpose, so such estimations would seem to be systematically unreliable by my lights. While such absolute probability estimations may sometimes be unreliable, it would be gratuitous

---

<sup>20</sup>For further discussion of Lipton’s suggestion, see (Dellsén, 2017b, 33-5).

to assume that this is systematically so. Thus, concludes the objection, IBE must be a reliable heuristic for absolute probabilities of explanatory hypotheses, contrary to what I have argued.

*Reply:* This objection is based on a misunderstanding of the argument of section 3.2. I did not argue that epistemic agents have no means of estimating the absolute probabilities of explanatory hypotheses. What I argued is that IBE cannot generally serve as a heuristic for this purpose, i.e. be an accessible decision procedure on the basis of which ordinary agents could assign rational credences to propositions. To say that IBE cannot play this role is not to say that it's impossible to reliably estimate absolute probabilities by some other means, e.g. by using some *other* heuristic. Indeed, nothing that I have said rules out that agents could estimate absolute probabilities directly, i.e. without using any heuristic at all. For example, the reason we think that Lamarck's theory is very likely false is hardly because it fails to possess some explanatory virtues in the appropriate quantity, but because the theory conflicts with observed evidence in straightforward ways. We don't need any explanatory heuristic to tell us that such a theory is very unlikely to be true, any more than we need an explanatory heuristic to tell us whether it is raining outside when we look out the window. It is important in this respect to recall that Bayesianism does not itself provide any heuristic for non-ideal agents to approximate its model of rational non-deductive reasoning. So my argument in section 3.2 leaves Bayesianism in no worse shape vis-à-vis reliable estimations of absolute probabilities than it was before the introduction of the heuristic conception of IBE.

Besides, as I discussed towards the end of section 3.2, I do think there are circumstances under which a kind of explanatory heuristic does allow us to estimate absolute probabilities, viz. when we have good reason to believe that all alternative explanations are so poor as to have a negligible influence on the marginal likelihood. Darwin's theory of natural selection would be a case in point in my view, since I think we have good reason to believe that all alternative explanations of biological evolution will provide very poor explanations of the evidence that has been accumulated over the past one and a half century. In my view, that is why we are entitled to say that Darwin's theory of natural selection is very likely true. So the argument of section 3.2

is not only compatible with agents making reliable estimations of absolute probabilities by other means than IBE; it is also compatible with agents using explanatory considerations in certain favorable circumstances to estimate that theories are very likely true.

## 4 CHOOSING WORKING HYPOTHESES BY IBE

The previous section argued that using IBE as a heuristic to absolute probabilities is deeply problematic in typical cases since we generally have considered only a fraction of the explanatory hypotheses that are relevant for estimating absolute probabilities by IBE. If this is correct, then IBE typically indicates only how the probability of one available explanatory hypothesis compares with that of another such hypothesis. However, since Bayesian reasoning (at least as traditionally conceived) operates with absolute as opposed to merely comparative probabilities, this might seem to preclude any substantive heuristic role for IBE in a Bayesian framework for non-deductive reasoning. This section argues that this conclusion would be too hasty: IBE can play an important heuristic role in Bayesian reasoning even though it is incapable of providing a guide to absolute probabilities in the traditional way. In short, I shall argue that through identifying the hypotheses with the highest probabilities in a given set of available hypotheses, IBE indicates which hypotheses Bayesian inquirers ought to make into the focal points of further testing and experimentation in the relevant domain.<sup>21</sup>

It is a platitude that scientists typically do not begin an investigation into an empirical question without having already formulated hypotheses about how the question might be answered. Such hypotheses will influence the course of their empirical research, e.g. by determining which things they choose to observe and which experiments they decide to carry out. In that sense, the

---

<sup>21</sup>There are similarities between the role I describe for IBE here and what Charles S. Peirce calls *Abduction*. In particular, Peirce describes Abduction as warranting only a probative kind of acceptance for the purposes of inductive examination (Kapitan, 1992, 12-17). However, in contrast to Peirce, I am not arguing here that IBE has a role to play in *generating* new hypotheses in what Reichenbach (1938) called “the context of discovery” (see, e.g., Minnameier, 2004; Campos, 2011).

gathering of scientific evidence is routinely guided by theoretical assumptions. The point applies in everyday cases as well. Consider, for example, our detective case: Before leaving the crime scene, the detective will have made observations, taken samples, and interviewed suspects. But where should she look? What should she sample? Who should she interview, and what questions should she ask? The answers to these questions will be determined by the detective’s assumptions about the case, including in particular her suspicion about the actual cause of the victim’s death. Generally, decisions about what evidence to gather will, at least for rational agents, be determined in large part by one’s theoretical assumptions and suspicions about the subject matter under investigation.

Of course, if one is logically and theoretically omniscient, as ideal Bayesian agents are, then one will have entertained the entire logical space of possible hypotheses and assigned probabilities to every single hypothesis in that space. So, for ideal Bayesian agents, decisions about what evidence to gather will be made on the basis of such probabilities directly. However, ordinary epistemic agents of the sort that IBE is meant to provide a heuristics for will seldom, if ever, be able to simultaneously entertain all hypotheses in logical space and assign a probability to each one. Ordinary agents of this sort will instead focus on a limited number – often only one – of the available hypotheses about the subject matter under investigation, which then serve as her primary guide to how she should proceed in the investigations that follow. I will refer to hypotheses that serve this role in an agent’s epistemic life as *working hypotheses*.<sup>22</sup>

Now, rational agents will clearly seek to minimize the risk of devoting time and resources to investigating hypotheses that further research will show to be false, i.e. dead-end hypotheses. Thus, all other things being equal, rational agents will seek to adopt working hypotheses that are maximally probable in light of the evidence. This is where I suggest that IBE enters the picture, for recall that through providing a heuristic for comparative probabilistic evalu-

---

<sup>22</sup>Note that here and throughout I use the plural “working hypotheses”. This is to emphasize that sometimes we may want to adopt more than one working hypotheses simultaneously, especially when no single hypotheses is decisively the most promising of those that are available.

ations, IBE in effect also provides a heuristic for ranking a given set of hypotheses in terms of their (objective Bayesian) probabilities. In particular, then, IBE can be used as a heuristic for identifying the hypotheses with the highest such probabilities in a given set of available hypotheses. My suggestion, then, is that IBE serves as a heuristic for identifying the most probable of the available explanatory hypotheses to be adopted as working hypotheses around which further inquiry will be structured.

To illustrate, consider once again our detective case. Since the detective does not want to waste her efforts on hypotheses that are unlikely to be corroborated by further investigation (and thus unlikely lead to conviction), she will seek to focus her current investigations on the most probable hypotheses available. Using explanatory considerations as a heuristic, she may plausibly come to conclude that  $H_b$  is by far the most probable hypothesis because of its superior explanatory qualities as compared with the available alternatives,  $H_s$  and  $H_a$ . Consequently, she adopts  $H_b$  as her only working hypothesis, which in turn will be manifested in the way in which she carries out further investigation. So, for example, the detective may interrogate the brother, gather his fingerprints, and check for gun powder on his sleeve, instead of (or at least prior to) gathering evidence on the victim's mental history, attempting to locate a suicide note, and so forth. Although the detective may not have a good estimation of the absolute (objective Bayesian) probabilities of the three hypotheses, she knows enough to make an informed decision which of the three hypotheses to pursue.

Notice that if IBE plays this rather modest role in an otherwise Bayesian framework for non-deductive reasoning, then it is entirely appropriate that the explanatory considerations to which IBE appeals provide a reliable heuristic only to comparative probabilities in the way I argued in the previous section. After all, IBE would still provide a guide to locating the most probable of the available hypotheses for the agent to adopt as working hypotheses, which enables the agent to minimize as far as possible the risk of adopting working hypotheses that turn out to be dead ends. While the risk of adopting only false hypotheses as working hypotheses may often be significant, an agent can do no better than to let the *most probable* available hypotheses guide her further investigations. Thus the choice of adopting the most probable available



hypotheses as a working hypotheses is made rational by the fact that doing so *optimizes* the agent's chances of adopting a true working hypothesis.<sup>23</sup>

I am suggesting that comparative probabilistic evaluations serve as the basis for rational decisions about which hypotheses to adopt as working hypotheses. One might worry about this on the grounds that, in general, comparative probabilities are not enough for rational decision according to standard Bayesian decision theory, which requires that the expected utility of each available course of action be calculated from absolute probabilities of their possible outcomes and the utilities of those outcomes. By recommending that decisions about which working hypotheses to adopt be made on the basis of comparative probabilities, it might seem as though I am contradicting this hugely successful and well-entrenched framework for rational decision making. It's important to see that this is not the case.

Just as heuristic conceptions of IBE generally do not aim to replace standard Bayesian epistemology, but rather to supplement the it by providing an accessible decision procedure to approximate Bayesian reasoning (see §2), my suggestion here is that comparative probabilities may provide an accessible heuristic to approximate standard Bayesian decision theory in the special case in which we are choosing between competing explanatory hypotheses. One of the reasons IBE can serve this role is that, other things being equal, the utility of choosing to adopt each such hypothesis when it is true/false can be assumed to be equal or at least comparable for truth-seeking agents such as scientists, since each *competing* explanatory hypothesis purports to explain the same range of phenomena as any other such hypothesis. For this reason, comparisons between the expected utility of choosing each available hypothesis as a working hypothesis will, other things being equal, depend only on their com-

---

<sup>23</sup>It may be worth highlighting that to adopt something as a working hypothesis is at least partly a *practical* choice as opposed to a purely *epistemic* one. Two agents,  $S_1$  and  $S_2$ , may have the exact same doxastic attitudes (i.e. beliefs or credences) towards a hypothesis  $H_i$  even though  $S_1$  has adopted  $H_i$  as a working hypothesis while  $S_2$  has not. The difference between  $S_1$  and  $S_2$  would be that  $S_1$  has decided to focus further investigations around  $H_i$  while  $S_2$  has not. This is a practical difference in that it involves a difference in the prospective actions of  $S_1$  and  $S_2$ . However, there is a sense in which the difference is also epistemic, since it concerns a difference in how  $S_1$  and  $S_2$  will carry out their respective epistemic investigations. For this reason, we may want to say that adopting something as a working hypothesis involves both practical and epistemic elements.

parative probabilities for truth-seeking agents. In sum, then, the suggestion I am making here does not conflict with standard Bayesian decision theory, but instead complements it by providing an accessible heuristic in the special case of choosing between competing explanatory hypotheses.

With that said, I want to acknowledge that there will be special cases in which this heuristic fails to guide us towards rational decisions about what working hypotheses to adopt. This is because, plausibly, even truth-seeking agents will sometimes be faced with a choice between competing explanatory hypotheses where the utilities of adopting each one when they are true/false are not uniform in this way. The clearest example of such a case will perhaps be where one of the hypotheses is more easily testable than the others. In that case, standard decision theory may dictate that it would be rational for truth-seeking agents adopt the most testable hypothesis as their working hypothesis even at the expense of other hypotheses that are more probable, since that may maximize expected utility in the long run. Another case of this kind might occur when one of the explanatory hypotheses combines with other hypotheses such as to help answer questions that its alternatives are unhelpfully silent on, in which case adopting the former might be rational by standard decision theory's lights even if it is not among the most probable competing explanatory hypotheses available. So, on this heuristic conception of IBE, there might well be cases in which IBE fails to serve its role as a guide for making rational decisions about what to adopt as one's working hypotheses.<sup>24</sup>

While this brings out a certain limitation of IBE on the current heuristic conception, it does not significantly undermine IBE's status as a heuristic for choosing which working hypotheses to adopt. After all, recall that any heuristic for approximating another form of reasoning (or, in this case, decision making) will provide only a rough and fallible guide to the latter, so it hardly counts much against the current conception of IBE as compared with alternative heuristic conceptions that there be cases in which it falters. Indeed, note that IBE still provides a reliable guide to expected utility maximizing regarding decisions about which available hypotheses to adopt as working hypotheses

---

<sup>24</sup>Thanks to a reviewer for this journal for pressing me on this issue.

*when other things are equal*. In particular, it still provides such a guide when the available competing explanatory hypotheses are (roughly) *equally* testable and informative on other questions, since the utility of adopting such hypotheses when they are true/false will certainly be (roughly) uniform. Furthermore, note that this heuristic may provide ordinary agents with useful information even in the problematic cases in which the available competing explanatory hypotheses are *not* roughly equal in these respects, since even in those cases it provides agents with information about how to internally rank those subsets of hypotheses that are equally testable and informative on other questions.<sup>25</sup>

## 5 CONCLUSION

The heuristic conception of IBE promises to show how IBE and Bayesianism are not only compatible (contra van Fraassen’s (1989) influential argument) but also complementary. I have argued, however, that there are limitations in principle to how much can be asked of IBE in this respect, since explanatory considerations are not typically suitable for indicating the absolute probability values with which Bayesianism is standardly seen as operating. In light of these limitations, I have argued that IBE is best construed as a heuristic for estimating which hypotheses have the highest probability among those explanatory hypotheses that are available at a given time. I argued that this helps ordinary agents identify which explanatory hypotheses to adopt as the working hypotheses around which further investigation is structured. On this view, IBE complements Bayesianism by providing a heuristic for deciding how to structure subsequent inquiry in a given domain so as to maximize the

---

<sup>25</sup>For example, suppose that in a set of available competing explanatory hypotheses  $\{H_1, \dots, H_n\}$ , some three hypotheses  $H_1$ ,  $H_2$ , and  $H_3$  are (roughly) equally testable and informative on other questions. In that case, our heuristic would provide a means to rank  $H_1$ ,  $H_2$ , and  $H_3$  such as to determine which of the three should be adopted as a working hypothesis *provided that any of them is* (and similarly for other subsets of equally testable/informative hypotheses in  $\{H_1, \dots, H_n\}$ ). In particular, if for example our heuristic ranks  $H_1$  above  $H_2$  and  $H_3$ , this rules out as irrational to adopt only  $H_2$  or  $H_3$  as working hypotheses but not also  $H_1$ .

likelihood of success.<sup>26</sup>

---

<sup>26</sup>This paper was presented at the UNC Chapel Hill Philosophy of Science Workshop in honor of Marc Lange (a.k.a. MarcFest) in November 2015. I am grateful to Marc and the rest of the audience, and especially my commentator Matt Kotzen. I am also indebted to Michael Bertrand and a reviewer for this journal for helpful written comments. Research for this paper was in part supported by an Irish Research Council New Horizons grant (Project ID: REPRO/2015/89).

## REFERENCES

- Bird, A. (2005). Abductive Knowledge and Holmesian Inference. In Gendler, T. S. and Hawthorne, J., editors, *Oxford Studies in Epistemology*, pages 1–31. Oxford University Press, Oxford.
- Bird, A. (2007). Inference to the Only Explanation. *Philosophy and Phenomenological Research*, 74:424–432.
- Campos, D. G. (2011). On the distinction between Peirce’s abduction and Lipton’s Inference to the Best Explanation. *Synthese*, 180:419–442.
- Chow, S. J. (2015). Many Meanings of “Heuristic”. *British Journal for the Philosophy of Science*, 66:977–1016.
- Dellsén, F. (2016a). Explanatory Rivals and the Ultimate Argument. *Theoria*, 82:217–237.
- Dellsén, F. (2016b). Realism and the Absence of Rivals. *Synthese*, DOI: 10.1007/s11229-016-1059-3.
- Dellsén, F. (2017a). Abductively robust inference. *Analysis*, DOI: 10.1093/analys/anx049.
- Dellsén, F. (2017b). Reactionary Responses to the Bad Lot Objection. *Studies in History and Philosophy of Science*, 61:32–40.
- Douven, I. (1999). Inference to the Best Explanation Made Coherent. *Philosophy of Science (Proceedings Supplement)*, 66:S424–S435.
- Dupre, J. (2002). The Lure of the Simplistic. *Philosophy of Science*, 69:284–293.
- Earman, J. (1992). *Bayes or Bust: A Critical Examination of Bayesian Confirmation Theory*. MIT Press, Cambridge, MA.
- Foster, J. (1982). Induction, Explanation, and Natural Necessity. *Proceedings of the Aristotelian Society*, 101:145–161.
- Good, J. I. (1971). 46656 varieties of bayesians. *American Statistician*, 25:62–63.

- Harman, G. (1965). The Inference to the Best Explanation. *The Philosophical Review*, 74:88–95.
- Harman, G. (1997). Pragmatism and reasons for belief. In Kulp, C. B., editor, *Realism/Antirealism and Epistemology*. Rowman and Littlefield, Lanham, MD. Reprinted in Harman 1999.
- Harman, G. (1999). *Reasoning, Meaning and Mind*. Oxford University Press, Oxford.
- Henderson, L. (2014). Bayesianism and Inference to the Best Explanation. *British Journal for the Philosophy of Science*, 65:687–715.
- Huemer, M. (2009). Explanationist Aid for the Theory of Inductive Logic. *British Journal for the Philosophy of Science*, 60:345–375.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.
- Kapitan, T. (1992). Peirce and the Autonomy of Abductive Reasoning. *Erkenntnis*, 37:1–26.
- Keil, F. C. (2006). Explanation and Understanding. *Annual Review of Psychology*, 57:227–254.
- Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan, London.
- Kvanvig, J. (1994). A Critique of van Fraassen’s Voluntaristic Epistemology. *Synthese*, 98:325–348.
- Laplace, P.-S. (1951). *A Philosophical Essay on Probabilities*. Dover, New York.
- Lewis, D. (1980). A Subjectivist’s Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, volume 2, pages 263–93. University of California Press, Berkeley.
- Lewis, D. (1986). Causal Explanation. In *Philosophical Papers, volume II*, pages 214–240. Oxford University Press, Oxford.
- Lipton, P. (1993). Is the Best Good Enough? *Proceedings of the Aristotelian Society*, 93:89–104.

- Lipton, P. (2004). *Inference to the Best Explanation*. Routledge, London, 2nd edition.
- Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, 10:464–470.
- Lombrozo, T. (2007). Simplicity and Probability in Causal Explanation. *Cognitive Psychology*, 55:232–257.
- Lombrozo, T. (2010). Explanation and Abductive Reasoning. In Holyoak, K. and Morrison, R., editors, *The Oxford Handbook of Thinking and Reasoning*, pages 260–276. Oxford University Press, Oxford.
- Lombrozo, T. (2011). The instrumental value of explanations. *Philosophy Compass*, 6:539–551.
- Lombrozo, T. (2016a). Explanation. In Sytsma, J. and Buckwalter, W., editors, *A Companion to Experimental Philosophy*, pages 491–503. Blackwell, Malden, MA.
- Lombrozo, T. (2016b). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences*, 20:748–759.
- Lycan, W. G. (1988). *Judgment and Justification*. Cambridge University Press, Cambridge.
- Lycan, W. G. (2012). Explanationist Rebuttals (Coherentism Defended Again). *The Southern Journal of Philosophy*, 50:5–20.
- McGrew, T. (2003). Confirmation, Heuristics, and Explanatory Reasoning. *British Journal for the Philosophy of Science*, 54:553–567.
- Minnameier, G. (2004). Peirce-Suit of Truth - Why Inference to the Best Explanation and Abduction Ought Not to Be Confused. *Erkenntnis*, 60:75–105.
- Niiniluoto, I. (1999). Defending Abduction. *Philosophy of Science (Proceedings Supplement)*, 66:S436–S451.
- Okasha, S. (2000). Van Fraassen’s Critique of Inference to the Best Explanation. *Studies in the History and Philosophy of Science*, 31:691–710.

- Pennington, N. and Hastie, R. (1992). Explaining the Evidence: Tests of the Story Model for Juror Decision Making. *Journal of Personality and Social Psychology*, 62:189–206.
- Preston, J. and Epley, N. (2005). Explanations Versus Applications: The Explanatory Power of Valuable Beliefs. *Psychological Science*, 10:826–832.
- Psillos, S. (2007). The Fine Structure of Inference to the Best Explanation. *Philosophy and Phenomenological Research*, 74:441–448.
- Read, S. J. and Marcus-Newhall, A. (1993). Explanatory Coherence in Social Explanations: A Parallel Distributed Processing Account. *Journal of Personality and Social Psychology*, 65:429–447.
- Reichenbach, H. (1938). *Experience and Prediction*. University of Chicago Press, Chicago.
- Rosenkrantz, R. D. (1977). *Inference, Model and Decision: Towards a Bayesian Philosophy of Science*. Dordrecht, Reidel.
- Skow, B. (2014). Are There Non-Causal Explanations (of Particular Events). *British Journal for the Philosophy of Science*, 65:445–467.
- Sloman, S. A. (1994). When Explanations Compete: The Role of Explanatory Coherence on Judgements of Likelihood. *Cognition*, 52:1–21.
- Sober, E. (1990). Contrastive empiricism. In Savage, W., editor, *Scientific Theories, Minnesota Studies in the Philosophy of Science*, vol. 14, pages 392–412. Minnesota University Press.
- Stanford, P. K. (2006). *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford University Press, Oxford.
- Thagard, P. (1989). Explanatory Coherence. *Behavioral and Brain Sciences*, 14:435–502.
- Thagard, P. R. (1978). The Best Explanation: Criteria for Theory Choice. *The Journal of Philosophy*, 75:76–92.
- Tversky, A. and Kahneman, D. (1984). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgement. *Psychological Review*, 91:293–315.



- van Fraassen, B. C. (1980). *The Scientific Image*. Clarendon, Oxford.
- van Fraassen, B. C. (1989). *Laws and Symmetry*. Clarendon Press, Oxford.
- Weintraub, R. (2013). Induction and Inference to the Best Explanation. *Philosophical Studies*, 166:203–216.
- Weisberg, J. (2009). Locating IBE in the Bayesian Framework. *Synthese*, 167:125–143.